United States Marine Corps
School of Advanced Warfighting
Marine Corps University
2076 South Street
Marine Corps Combat Development Command
Quantico, Virginia 22134-5068

MASTER OF OPERATIONAL STUDIES

TITLE:

COMPUTATIONAL LINGUISTICS IN
MILITARY OPERATIONS

AUTHOR:

MARC U. CROPP
LIEUTENANT COLONEL (GS), GERMAN ARMY

AY 09-10

Mentor: ____
Approved: ______________________________________
Date: ____

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2010**	2. REPORT TYPE	3. DATES COVERED **00-00-2010 to 00-00-2010**
4. TITLE AND SUBTITLE **Computational Linguistics in Military Operations**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Marine Corps University,School of Advanced Warfighting,2076 South Street,Quanitico,VA,22134-5068**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Thesis. QUOTATION FROM, ABSTRACTION FROM, OR REPRODUCTION OF ALL OR ANY PART OF THIS DOCUMENT IS PERMITTED PROVIDED PROPER ACKNOWLEDGEMENT IS MADE.

14. ABSTRACT
Computational linguistics can significantly enhance battlespace awareness and support information dominance at the operational and tactical level of war in future warfare. Discussion: Mastering culture and language in a foreign country is decisive to understand the operational environment. In addition, the ability to understand and speak a foreign language is a prerequisite to achieve truly comprehension of an unfamiliar culture. Lasting operations in Afghanistan and Iraq and the necessity to breach the language gap lead to progress in the field of Machine Translation and the development of technical solutions to close the gap in the past decade. This paper examines the current development and evaluates the strength and weaknesses of present Machine Translation. Automated language processing comprising foreign to English translation, automatic speech to text transcription, and information management and text processing is a way to mitigate the complexity to enhance battlespace awareness with current available systems. However, the only way to achieve a breakthrough in translation technology is to decode the DNA of a language. Decoding a language and process it automatically is the task of Computational Linguistics. Conclusion: The current developments in the field of Machine Translation driven by enduring military operations and of the shelf solutions are a way to mitigate the existing language gap. However, fundamental progress can only be achieved by basic research in the field of Computational Linguistics.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Public Release**	**27**	

Executive Summary

Titel: Computational Linguistics in Military Operations

Author: Lieutenant Colonel (GS) Marc U. Cropp, German Army

Thesis: Computational linguistics can significantly enhance battlespace awareness and support information dominance at the operational and tactical level of war in future warfare.

Discussion: Mastering culture and language in a foreign country is decisive to understand the operational environment. In addition, the ability to understand and speak a foreign language is a prerequisite to achieve truly comprehension of an unfamiliar culture. Lasting operations in Afghanistan and Iraq and the necessity to breach the language gap lead to progress in the field of Machine Translation and the development of technical solutions to close the gap in the past decade. This paper examines the current development and evaluates the strength and weaknesses of present Machine Translation. Automated language processing comprising foreign to English translation, automatic speech to text transcription, and information management and text processing is a way to mitigate the complexity to enhance battlespace awareness with current available systems. However, the only way to achieve a breakthrough in translation technology is to decode the DNA of a language. Decoding a language and process it automatically is the task of Computational Linguistics.

Conclusion: The current developments in the field of Machine Translation driven by enduring military operations and of the shelf solutions are a way to mitigate the existing language gap. However, fundamental progress can only be achieved by basic research in the field of Computational Linguistics.

DISCLAIMER

THE OPINIONS AND CONCLUSIONS EXPRESSED HEREIN ARE THOSE OF THE INDIVIDUAL STUDENT AUTHOR AND DO NOT NECESSARILY REPRESENT THE VIEWS OF EITHER THE MARINE CORPS COMMAND AND STAFF COLLEGE OR ANY OTHER GOVERNMENTAL AGENCY. REFERENCES TO THIS STUDY SHOULD INCLUDE THE FOREGOING STATEMENT.

QUOTATION FROM, ABSTRACTION FROM, OR REPRODUCTION OF ALL OR ANY PART OF THIS DOCUMENT IS PERMITTED PROVIDED PROPER ACKNOWLEDGEMENT IS MADE.

Table of Contents

"At the heart, war will always involve a battle between two creative human forces. Our enemies are always learning and adapting. They will not approach conflicts with conceptions or understanding similar to ours."

The Joint Operational Environment, 2008

Introduction

The current and future operational environment is characterized by structural and dynamic complexity. This complexity is determined by a variety of different dimensions. One of these dimensions is a seemingly ever present language barrier between the operating forces and the opposing forces, belligerents, and the host nation population. This same can exist within a coalition as well.[1] Closing the language barrier not only reduces complexity, it can serve to mitigate other dimensions of complexity, like foreign culture or insights into an adversaries way of thinking. The one who masters a broad variety of languages in depth can gain a distinct operational advantage.

For military organizations it is a matter of vital importance to provide the necessary capabilities for successful task accomplishment regardless of complexity. Breaching the language gap is one challenge the military has to cover in order being prepared for future operations. The U.S. Marine Corps, like other services, emphasized the necessity for military personnel to learn foreign languages.[2] However, personal limitations, time constraints, and operational requirements limit the progress in closing the language gap. In addition, changing the operational environment generally requires a different set of language skills.

Therefore, the education of military personal requires more than one line of approach to close the language gap. A true step ahead is to automate the process of translation. Computational Linguistics (CL) and within this field of science Machine Translation (MT), provides solutions for the call of timely and on the spot available translation. Hence, *Computational Linguistics can significantly enhance battlespace awareness and supports information dominance at the operational and tactical level of war in future warfare*.

Definitions

Computational Linguistics – “the interdisciplinary field which involves both linguistics and computer science, and is concerned with (1) automatising the analysis of text and speech corpora and (2) developing precise models of grammars and lexica which can be processed automatically.“[3] Hence, Computational Linguistics is the theoretical foundation of Machine Translation. Progress in this field of science can be compared with decoding the human Deoxyribonucleic acid (DNA.)

Battlespace awareness is defined as “knowledge and understanding of the operational area's environment, factors, and conditions, to include the status of friendly and adversary forces, neutrals and noncombatants, weather and terrain, that enables timely, relevant, comprehensive, and accurate assessments, in order to successfully apply combat power, protect the force, and/or complete the mission.”[4] The degree of battlespace awareness is tightly connected to the access and evaluation of information. In foreign environments the degree of battlespace awareness depends on access to and translation of written or oral information. The translation has to be timely, comprehensive, and accurate. Otherwise the assessment of the information will not meet the necessary quality to enhance the degree of battlespace awareness, in contrary; the assessment can lead to disastrous decisions if the content is not translated in an appropriate way.

Information superiority is the condition for information dominance and closely related to battlespace awareness. Information superiority is defined as the “capability to collect, process, and disseminate an uninterrupted flow of information while exploiting or denying an adversary's ability to do the same.”[5] The ability to achieve information superiority is in the same way linked to the ability to translate foreign languages as the degree of battlespace awareness described before.

Information dominance is "the degree of information superiority that allows the possessor to use information systems and capabilities to achieve an operational advantage in a conflict or to control the situation in operations other than war while denying those capabilities to the adversary."[6] In other words, by gaining knowledge advantage over an adversary a friendly commander achieves information dominance. Translation capability and quality set therefore also the precondition for information dominance.

Translation Methods

The required methods of translation for military purpose are diverse. At higher headquarters on the operational level the requirement for text-to-text-translation (T2T) and speech-to-text-translation (S2T) primarily exists. The purpose is to translate foreign documents, the content of foreign websites on the internet, and foreign broadcasts on the television or radio to broaden the base of information. Additionally, on the tactical level the requirement for speech-to-speech-translation (S2S) grows. S2S enables the communication with the local foreign population and is therefore essential for any kind of military operations on foreign soil. At present, military forces remain heavily depend on linguists to accomplish the translation requirements in all three methods.

Foreign language speech and text are an indispensable source of intelligence. However, the vast majority available is still unexamined. Foreign language data and their corresponding providers are massive and growing in numbers daily. Moreover, because the time to transcribe and translate foreign documents is labor intensive, compounded by the lack of linguists with suitable language skills to review it all, much foreign language speech and text are not exploited for intelligence purposes in order to enhance battlespace awareness or gain information dominance. At present and in future it is and it will be impossible to find, train, or pay enough people to accomplish this task. New and powerful foreign language technology is needed to allow English-speaking analysts to exploit and understand vastly more foreign speech and text than is currently possible.[7]

Evaluation of Present Machine Translation

The capacity of human and capability of machine translation within the military is limited at present. During 2009, U.S. Forces in Iraq and Afghanistan contracted about 12.000 host nation linguists.[8] However, an automated translation process can outperform these host nation linguists not only in terms of capacity but also in several other aspects:

Strength of Machine Translation

(1) Credibility – Ability of a language translation system to provide credible, not intentionally misleading, two-way translation of voice and text.

Host nation linguists, are frequently employed locally and are the most plentiful resource pool but their credibility is rated inadequately. As local nationals, these host nation linguists first loyalty is most likely to the host nation or ethnic group and frequently not to the U.S. military. Some linguists may have hidden motives or a concealed agenda for political or personal reasons. Therefore, the types of information host nation linguists can overhear are limited. Unlike human linguist, MT systems have no potential for bias or hidden agenda. Hence, they can be evaluated as being highly credible.[9]

(2) Deployability – Ability to deploy a language system to support all missions when and where language translation capabilities are required within a specified time frame

Compared to MT system, acquisition of host nation linguists required long lead time as the contractor cannot begin the local hiring process until there is a stable and permissive environment. This leads to its lower rating in deployability.[10] In addition, from the study conducted by Battelle Memorial Institute, Commanders agreed that in many cases, contract linguists are able to hold their units hostage and offered the following comments about contract linguists:[11]

(A) They refuse to support certain missions with little or no consequence.

(B) The contractor responsible for contract linguist management is seldom seen.

(C) Many contract linguists are physically unable to operate at the required operational tempo.

The above sheds light on the problems associated with deployment of host nation linguists. On the other hand, MT systems are readily available for deployment so long as the units are assigned the required number of MT systems with the appropriate language modules and mission sets to support their missions. MT systems also have an added advantage over host nation linguists who are at risk of being targeted by adversary during deployment to the area of operation as well as after the conducted mission.

(3) Translation requirement fill – Ability of language translation solutions to satisfy tasks with large number of “linguistic points of presence”

MT systems provide the capability to meet the requirements when there were large numbers of “linguistic points of presence,” defined as points in space where speech and/or text translation support is required. With limited number of linguists assigned to the units, host nation linguists comparatively fared poorly in this aspect. In addition, most of the military operations require linguist teams to be able to support 24 hour operations, so a minimum of four linguists per team is necessary. This aggregates the problem of limited number of linguists to meet translation requirements both in space and time.[12]

(4) Translation speed – Number of words per minutes that a T2T, S2T, T2S, or S2S system is capable of translating

The primary advantage of MT systems is translation speed. Fast translation speed could lead to operational advantages. The translations speed for an average human, whether S2S or T2T, is slow. S2S translations will take place at less than a conversational pace. The average human translator can translate approximately 30 to 60 words of text per minute.[13] The MT T2T translation capability is significantly faster than that of host nation linguist, though at present the translations are much less precise on anything above Interagency Language Roundtable (ILR) level 2.[14] For example, the currently available Documentation Exploitation (DOCEX) system is able to distill useful intelligence from multilingual sources eight to ten times faster than traditional manual methods, thereby enabling the Intelligence units to focus their limited linguistic resources on documents that have the highest probability of containing value.[15]

(5) Consistency – Ability of a language translation system to give consistent translation

MT systems have a better memory that is unmatched by human translators. It can store translated documents and re-use phrases that have already been translated, resulting in highly consistent translation throughout missions.[16] Provided that MT systems give an accurate translation, consistent translation is certainly desirable.

However, there are indeed several limitations within MT at present. These limitations will be determined in the next section.

Current Weakness of Machine Translation

Currently, the capability of automated translation is limited by several factors. A number of key factors are listed and described in the following.

(1) Translation level capability – Ability of a language translation system to render consistent two-way translations at a level based upon the ILR description

The Lincoln Laboratory at the Massachusetts Institute of Technology (MIT) in cooperation with the Department of Brain and Cognitive Sciences and Defense Language Institute Foreign Language Centre (DLIFLC) conducted an experiment designed to measure human readability of machine generated text. This three part experiment focused on S2T and T2T translation. The results of their experiment showed that the current state-of-the-art MT technologies can achieve an ILR score of between 1+ to 2 in S2T and 2 to 2+ in T2T translation. These results indicated that MT systems have the capability to accomplish vast majority of tasks with low level translation requirement, at the ILR level 2 or less. On the other hand, those host nation linguists who possess the required linguistic ability in English have the potential to achieve an unmatched high ILR score of 5, which is high enough to meet any translation requirement.[17]

(2) Extensibility – Ability of a translation system to add additional language modules

It is impossible for one-fit-all solution, so MT systems are designed for selected language pairs within certain domains. The process to add new languages to a MT system takes time and the timeline for developing a new language is similar to that of training a new linguist.[18] Hence, current MT systems are unable to meet time sensitive translation requirements that call for development of a new language. Therefore, at present host nation linguists have an advantage over MT systems and even military linguists for contingency operations. Operation Joint Endeavor (OJE), the initial peacekeeping operation in Bosnia-Herzegovina, began in December

1995. Prior to that mission, the Army had very little need for Serbian-Croatian linguists, and was caught unprepared for the large requirement of OJE. Though, the U.S. Army Europe (USAEUR) linguist support contract enabled the Army to acquire approximately 500 Serbian-Croatian linguists in a relatively short amount of time.[19]

(3) Versatility – Ability of a translation system to deal adequately with various complexities of language

One of the biggest limitations of MT systems today is their inability to deal adequately with the various complexities of language that humans handle naturally: ambiguity, syntactic irregularity, multiple word meanings and the influence of context.[20] A classic example is illustrated in the following pair of sentences: "Time flies like an arrow" and "Fruit flies like an apple". A computer can be programmed to understand either of these examples, but not to distinguish between them. A computer translation is similar to a translation done by a human without a deep knowledge of the target language.[21]

Alan Melby, professor of linguistics at Brigham Young University, points out that "Being a native or near-native speaker involves more than just memorizing lots of facts about words. It includes having an understanding of the culture that is mixed with the language. It also includes an ability to deal with new situations appropriately. No dictionary can contain all the solutions since the problem is always changing as people use words in unusual ways."[22]

Improvement for the Future

To enhance battlespace awareness and support information dominance the way forward has to follow two directions. First, automated language processing has to be improved. This will support primarily the T2T and S2T capability and therefore, the operational level of war. The second path will lead into the core of computational linguistics. The research in low- and middle-density languages has be improved in order to enable high quality S2S translation. This is essential to support MT solutions for communication of individuals with diverse language background and therefore for the tactical level. While the solution for the first path seems to be attainable in the near future, the second path appears to be much longer and will require more time.

Automated Language Processing

Three technologies determine automated language processing and will realize significant improvement. These are: (1) foreign-to-English translation technologies, (2) speech-to-text transcription technologies, and (3) information management and text processing technologies (also applicable for the contextual exploitation capability). Improvements in these technologies should allow automated processes and English-speaking users to examine and analyze all multilingual speech and text that is available in the information space; allow any user—be it primarily an operational and strategic planner; analyst; or decision-maker—to acquire basic language proficiency in days and expert language proficiency in months, for any language; and to continue improvements in word error rate, precision and recall, and usability measures, such as effectiveness, efficiency, and user satisfaction.[23]

One example of an R&D program in this area that integrates all three constituent technologies is the Global Autonomous Language Exploitation (GALE) program of the Defense Advanced Research Project Agency (DARPA). The GALE program is developing and applying computer software technologies to absorb, analyze, and interpret huge volumes of speech and text in multiple languages, eliminating the need for linguists and analysts. It is also developing the ability to automatically provide relevant, distilled actionable information to military command and personnel in a timely fashion. Automatic processing “engines” convert and distill the data, delivering pertinent, consolidated information in easy-to-understand forms to military personnel and monolingual English-speaking analysts in response to direct or implicit requests.[24]

Foreign-to-English Translation

Goals for foreign-to-English translation include: (1) providing high accuracy machine translation and structural metadata annotation from multilingual text document and speech transcription input at all stages of processing and across multiple genres, topics, and mediums (such as, Arabic, Chinese, the Web, news, blogs, signals intelligence, and databases); (2) understanding—or at least deriving semantic intent from—input strings regardless of source; (3) reconciling and resolving semantic differences, duplications, inconsistencies, and ambiguities across words, passages, and documents; (4) more efficient discovery of important documents, more relevant and accurate facts while decreasing the amount of time required to do it, and passages for distillation; (5) providing enriched translation output that is formatted, cleaned-up, clear, unambiguous, and meaningful to decision-makers; (6) eliminating the need for human intervention and minimized delay of information delivery; and (7) fast development of new language capability, swift response to breaking events, and increased portability across languages, sources, and information needs. Some examples of critical contributing technologies include: improved dynamic language modeling with adaptive learning; advanced machine translation technology that utilizes heterogeneous knowledge sources; better inference models; better tagging and annotation algorithms; language-independent approaches to create rapid, robust technology that can be ported cheaply and easily to any language and domain; syntactic and semantic representation techniques to deal with ambiguous meaning and information overload; and cross- and monolingual, language-independent information retrieval to detect and discover the exact data in any language quickly and accurately, and to flag new data that may be of interest.[25]

Automatic Speech-to-Text Transcription

Automatic speech-to-text transcription seeks to produce rich, readable transcripts of foreign news broadcasts and conversations (over noisy channels and/or in noisy environments) despite widely-varying pronunciations, speaking styles, and subject matter. In general, the two basic components of rich transcription are S2T conversion (finding and transcribing relevant words) and metadata extraction (pulling out features to annotate the transcripts to provide more useful information to the user).[26] There are also two basic approaches to S2T transcription—those that use constrained vocabularies (such as, Phraselator), and those that do not. Recent achievements (2004) include word error rates of 26.3 percent and 19.1 percent at processing speeds of 7 and 8 times slower than real-time on Arabic and Chinese news broadcasts.[27] Goals for S2T transcription include: (1) providing high accuracy multilingual word-level transcription from speech at all stages of processing and across multiple genres, topics, speakers, and channels (such as Arabic, Chinese, and other relevant speech dialects from news broadcasts, talk shows, the Web, signals intelligence, and databases); (2) representing and extracting "meaning" out of spoken language by reconciling and resolving jargon, slang, code-speak, and language ambiguities; (3) dynamically adapting to (noisy) acoustics, speakers, topics, new names, speaking-styles, and dialects; (4) improving relevance to deliver the information decision-makers need; (5) assimilating and integrating speech across multiple sources to support exploration and analysis to enable natural queries and drill-down; and (6) increased portability across languages, sources, and information needs.[28] Some examples of critical contributing technologies include: improved acoustic modeling; robust feature extraction; better discriminative estimation models; improved language and pronunciation modeling; and language independent approaches that are able to learn from examples by using algorithms that exploit advances in computational power plus the large quantities of electronic speech and text that are now available. The ultimate goal is to create rapid, robust technology that can be ported cheaply and easily to other languages and domains.

Information Management and Text Processing

There are many technologies that fall within the category of information management and text processing; too many to address in detail here. Some key technologies of particular value are:

Information retrieval has been responsible for the development of many useful algorithms and techniques for document analysis. This is in part due to the statistical nature of information retrieval, which itself derives from the vast amount of data such programs typically face. The essential problems in information retrieval are concerned with both similarity and ranking. Binding similar documents together makes information retrieval conceptually coherent; ranking them in order of relevancy to a query makes it efficient.[29]

"Advanced search" uses a combination of an advanced keyword approach (to compensate for common typing/spelling confusions and idiosyncrasies) and probabilistic latent semantic analysis to ascertain if a particular topic is being discussed without using specific keywords.[30]

Latent semantic analysis is one of a large class of unsupervised machine learning techniques that transform the original representation of texts to a new representation reflecting patterns of word occurrences in a large corpus of texts. In some situations, using this new representation can provide a small improvement in the effectiveness of processes such as search or classification applied to the text versus using a representation based on the original words and phrases of the document. Latent semantic analysis is mostly likely to provide an advantage when the data has an underlying structure (modeled as dimensions in a real-valued space) that matches up nicely with the categories to which a system is trying to assign texts.[31]

Entity extraction methods extract key facts from documents by accurately mining information from free text based on user requirements. These approaches were developed to be most effective when formal reports and articles are the materials for analysis. Entity extraction techniques are likely to be less effective in the chat medium, where content is less structured and language use is less formal. Abbreviations, misspellings, slang, and more speech-like constructions are the norm rather than the exception in chat. Although name translation remains problematic, automatic name extraction (or tagging) works reasonably well in English, Chinese, and Arabic. Researchers increasingly focus on sophisticated techniques for extracting information about entities, relationships, and events.[32]

Relationship extraction is much harder than entity extraction, and is important when seeking to extract entities and their relationships from textual narratives about activities, people, materials, and organizations, for example. Advanced techniques are able to efficiently and accurately discover, extract, and link sparse evidence contained in large amounts of unclassified and classified data sources such as public news broadcasts or classified intelligence reports.[33]

Detection uses advanced techniques to detect and discover the exact information a user seeks quickly and effectively and to flag new information that may be of interest. Cross-language information retrieval is the current focus of the research community with recent results showing the technique can work roughly as well as monolingual retrieval.[34]

Summarization reduces (substantially) the amount of text that people have to read. Researchers are now working on techniques for automatic headline generation (for single documents) and for multi-document summaries (of clusters of related documents).[35]

Graphical representations are critical to enable “connecting the dots” when representing data and patterns as graphs. Patterns specified as graphs with nodes representing entities such as people, places, things, and events; edges representing meaningful relationships between entities; and attribute labels amplifying the entities and their connecting links, are matched to data represented in the same graphical form. These highly connected evidence and pattern graphs also play a crucial role in constraining the combinatorics of the iterative graph processing algorithms such as directed search, matching, and hypothesis evaluation.[36]

Link discovery starts from known entities and uses statistical, knowledge-based, and graph-theoretic techniques to identify explicit links, infer implicit links, and evaluate their significance. Search is constrained by expanding and evaluating partial matches from known starting points, rather than the alternative of considering all possible combinations. The high probability that linked entities will have similar class labels can be used to increase classification accuracy.[37]

Pattern learning techniques can induce a pattern description from a set of exemplars. Such pattern descriptions can assist an analyst in discovering unknown terrorist activities in data. These patterns can then be evaluated and refined before being considered for use in detecting potential terrorist activity. Pattern learning techniques are also useful in enabling adaptation to changes in terrorist behavior over time.[38]

The military has to sustain a long-term commitment and robust effort to develop and adapt automated language processing technologies. This effort has to involve tapping into and leveraging commercial research and development work and investments. However, it also requires focused investments for those particular languages and dialects which the military uniquely require.

The Vision

Provide soldiers the capability to listen instantly to foreign speech and to communicate with people in a foreign language as if they had advanced linguistic abilities equivalent with a high ILR score of 5. This will breach the language gap for every Marine and soldier. Hence, the capability of every individual in the services will be increased and the ability to understand the operational environment improved. Military leaders will be able to communicate their ideas instantly and are able to take the feedback simultaneously without any filter. In the next step squads will record the communication of the civil population during patrolling and will be able to assess the taped to gather further information. Foreign languages will no longer cast a cloud over battlespace awareness.

Conclusions

The path to achieving immediate automated translation is still long. However, the benefit to enhance battlefield awareness and achieve information dominance is worth the endeavor. Future development will depend on the progress in Computational Linguistics. The progress achieved in this specific field of science is the foundation for future development in Machine Translation and it is therefore the prerequisite for further development to enhance battlespace awareness and support information dominance.

Endnotes

[1] Northam, Jackie. *Pentagon Ramps Up Training to Cut Language Gap*. National Public Radio. February 9, 2007. <http://www.npr.org/templates/story/story.php?storyId=7299421> (accessed: December 12, 2009)
[2] U. S. Marine Corps. *Marine Corps - Vision and Strategy 2025*. PCN 50100654800, p 20
[3] Webpage University Bielefeld. <http://wwwhomes.uni-bielefeld.de/~gibbon/Classes/Classes2005WS/HTMD/classnotes.html> (accessed: December 12, 2009)
[4] U. S. Department of Defense. *Dictionary of Military and Associated Terms*. US Department of Defense. October 31, 2009. <http://dtic.mil/doctrine/dod_dictionary/data/b/18268.html> (accessed: December 12, 2009)
[5] U. S. Army. *Army Vision 2010*. <http://www.army.mil/2010/information_superiority.htm> (accessed: December 12, 2009)
[6] Headquarters Department of the Army. *FM 100-6, Information Operations*. August 1996, p 1-9
[7] Office of the Under Secretary of Defense for Acquisition Technology and Logistics Washington DC. *Defense Science Board 2006 Summer Study on 21st Century Strategic Technology Vectors, Volume 2: Critical Capabilities and Enabling Technologies*. Ft. Belvoir: Defense Technical Information Center, 2007. <http://handle.dtic.mil/100.2/ADA464370> (accessed: December 12, 2009), p 33
[8] Schwartz, Mosche. Department of Defense Contractors in Iraq and Afghanistan: Background and Analysis. Congressional Research Service. December 2009. <http://www.fas.org/sgp/crs/natsec/R40764.pdf> (accessed: December 28, 2009), p 9-13
[9] Headquarters Department of the Army. *FM 3-24 Counter Insurgency*. December 2006. <https://atiam.train.army.mil/soldierPortal/atia/adlsc/view/public/23285-1/FM/3-24/FM3_24.pdf> (accessed: December 12, 2009), p C-9
[10] Laganosky, Vaughn. United States Army Machine Foreign Language Translation Requirements and Capabilities. Machine Translation Summit, Ottawa Canada, 2009. <http://www.mt-archive.info/MTS-2009-Lagnosky-ppt.pdf> (accessed: December 12, 2009)
[11] Hampson E., Jones T., Strictland M.. *Determination of the Army's Requirement for a Cadre of Professional Linguists at Level 3/3/3*. Battelle Memorial Institute 2007.
[12] Ibid.
[13] Cerutti, Edward A., Tim Hunter. *Sequoyah Foreign Language Translation System Analysis of Alternatives (AoA) Final Report*. 2005
[14] The scale is from 1 to 5; 5 is the highest level.
[15] CACI Ever Vigilant. *Document and Records Management – Document Exploitation*. <http://www.caci.com/fcc/dikm/doc_manage_docex.shtml> (accessed: December 12, 2009)
[16] Goutte, Cyril. *Learning Machine Translation*. Cambridge, Mass: MIT Press, 2009, p 3
[17] Cerutti, Edward A., Tim Hunter. *Sequoyah Foreign Language Translation System Analysis of Alternatives (AoA) Final Report*. 2005
[18] Ibid.
[19] Hampson E., Jones T., Strictland M. *Determination of the Army's Requirement for a Cadre of Professional Linguists at Level 3/3/3*. Battelle Memorial Institute 2007.
[20] Goutte, Cyril. *Learning Machine Translation*. Cambridge, Mass: MIT Press, 2009, p 244
[21] DiploFoundation. *Machine Translation*. <http://www.diplomacy.edu/Language/Translation/machine.htm> (accessed: December 12, 2009)
[22] Ibid.
[23] Mac Pherson, Cecil, Devin Rollis, Irene Zehmisch. *Meeting Army Foreign Language Requirements with the Aid of Machine Translation*. <www.mt-archive.info/AMTA-2008-MacPherson.pdf> (accessed: December 12, 2009), p 435
[24] Information Processing Techniques Office. *Global Autonomous Language Exploitation (GALE)*. <http://www.darpa.mil/ipto/Programs/gale/gale_approach.asp> (accessed: December 12, 2009). GALE program Graphic in the Appendix A.
[25] Mac Pherson, p 436
[26] Ibid, p 348
[27] Popp, Robert. *Countering Terrorism Trough Information Technology*. Communications of the ACM. March 2004. Vol. 47/3. <http://information-retrieval.info/taipale/papers/p36-popp.pdf> (accessed: December 12, 2009) p 42

[28] Mac Pherson, p 438
[29] Manning, Christopher. *An Introduction to Information Retrieval*. Cambridge University Press, 2009. <http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf> (accessed: December 12, 2009), p 3
[30] Hofmann, Thomas. *Probabilistic Latent Semantic Analysis To appear in: Uncertainity in Artificial Intelligence*, UAI'99, Stockholm, p 1
[31] Shapiro, Amy. *The use of Latent Semantic Analysis as a Tool for the Quantitative Assessment of Understanding and Knowledge.* <https://www.msu.edu/~kidwaikh/cep901b/literaturereview/week04lsa.htm> (accessed: December 12, 2009)
[32] Encyclopedia Britannica. <http://www.britannica.com/bps/additionalcontent/18/35789701/Chapter-3-Entity-Extraction-Statistical-Methods> (accessed: December 12, 2009)
[33] Encyclopedia Britannica. <http://www.britannica.com/bps/additionalcontent/18/35789702/Chapter-4-Relationship-Extraction> (accessed: December 12, 2009)
[34] Diekema, Anne R., Wen-Yuan, Hsiao. Cross-*Language Information Retrieval using Dutch Query Translation*. Syracuse University, School of Information Studies, 2001.
[35] CL Research. *Knowledge Management System*. <http://www.clres.com/kms.html> (accessed: December 12, 2009)
[36] RAND. “Connecting the Dots” in Intelligence. RAND Corporation, Santa Monica, CA, 2005. <http://www.rand.org/pubs/research_briefs/RB9079/RAND_RB9079.pdf> (accessed: December 12, 2009)
[37] Mooney, Raymond J., Prem, Melvill, Lappoon Rupert, Tang, Jude, Shavlik, Inês, de Castro Dutra, David, Page, Vitor Santos, Costa. *Relational Data Mining with Inductive Logic Programming for Link Discovery*. Department of Computer Sciences, University of Texas, 2005.
[38]Riloff, Ellen, Jancye Wiebe. *Learning Extraction Patterns for Subjective Expressions*. School of Computing University of Utah, 2003 <http://www.cs.utah.edu/~riloff/pdfs/emnlp03.pdf> (accessed: December 12, 2009)

Bibliography

Anderson, Timothy, et al. *Implications of Multilingual Interoperability of Speech Technology for Military Use (Les Implications De L'interoperabilite Multilingue Des Technologies Vocales Pour Applications Militaires)*. Ft. Belvoir: Defense Technical Information Center, 2004. <http://handle.dtic.mil/100.2/ADA431428>.

Association for Computational Linguistics. *Computational Linguistics*. Association for Computational Linguistics, 2008.

, and Heinz Josef Weber. *sentation und Textbezug = New approaches in machine translation : knowledge representation and discourse models* bingen: M. Niemeyer, 1986.

Biermann, Joachim, Louis de Chantal, Reinert Korsnes, Jean Rohmer, and Cagatay Uendeger. *From Unstructured to Structured Information in Military Intelligence - Some Steps to Improve Information Fusion*. Ft. Belvoir: Defense Technical Information Center, 2004. <http://handle.dtic.mil/100.2/ADA460220>.

Cerutti, Edward A., Tim Hunter. *Sequoyah Foreign Language Translation System Analysis of Alternatives (AoA) Final Report*. 2005.

Chambers, James J , Jr. *Force Protection for the 21st Century Expeditionary Aerospace Forces: Security Forces Need for Organic Ground Intelligence*. Ft. Belvoir: Defense Technical Information Center, 1999. <http://handle.dtic.mil/100.2/ADA395785>.

CL Research. *Knowledge Management System*. <http://www.clres.com/kms.html>.

Degen, Edmund J. *Dominant Land Forces for 21st Century Warfare*. Ft. Belvoir: Defense Technical Information Center, 2007. <http://handle.dtic.mil/100.2/ADA487296>.

DiploFoundation. *Machine Translation*. <http://www.diplomacy.edu/Language/Translation/machine.htm>.

Goutte, Cyril. *Learning Machine Translation*. Cambridge, Mass: MIT Press, 2009.

Gray, Colin S. *Another Bloody Century: Future Warfare*. London: Phoenix, 2006.

Hampson E., Jones T., Strictland M.. *Determination of the Army's Requirement for a Cadre of Professional Linguists at Level 3/3/3*. Battelle Memorial Institute 2007.

Haney, Eric L., and Brian Thomsen. *Beyond Shock and Awe: Warfare in the 21st Century*. New York: Berkley Caliber, 2006.

Hauenschild, Christa. *Machine translation and translation theory*. Berlin [u.a.]: Mouton de Gruyter, 1997.

Hausser, Roland. *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Berlin [u.a.]: Springer, 2001.

Headquarters Department of the Army. *FM 100-6, Information Operations*. August 1996.

Headquarters Department of the Army. *FM 3-24 Counter Insurgency*. December 2006. <https://atiam.train.army.mil/soldierPortal/atia/adlsc/view/public/23285-1/FM/3-24/FM3_24.pdf>.

Information Processing Techniques Office. *Global Autonomous Language Exploitation (GALE)*. <http://www.darpa.mil/ipto/Programs/gale/gale_approach.asp>.

Jurafsky, Daniel S., and James H. Martin. *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, N.J.: Prentice-Hall, 2009.

Koh, James. *Machine Translation: A Key to Information Supremacy and Knowledge-Based Operations*. USAWC strategy research project. Carlisle Barracks, PA: U.S. Army War College, 2003.

Laganosky, Vaughn. United States Army Machine Foreign Language Translation Requirements and Capabilities. Machine Translation Summit, Ottawa Canada, 2009. <http://www.mt-archive.info/MTS-2009-Lagnosky-ppt.pdf>.

Mac Pherson, Cecil, Devin Rollis, Irene Zehmisch. *Meeting Army Foreign Language Requirements with the Aid of Machine Translation*. <www.mt-archive.info/AMTA-2008-MacPherson.pdf>.

Manning, Christopher. *An Introduction to Information Retrieval*. Cambridge University Press, 2009. <http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>.

Mooney, Raymond J., Prem, Melvill, Lappoon Rupert, Tang, Jude, Shavlik, Inês, de Castro Dutra, David, Page, Vitor Santos, Costa. *Relational Data Mining with Inductive Logic Programming for Link Discovery*. Department of Computer Sciences, University of Texas, 2005.

Riloff, Ellen, Jancye Wiebe. *Learning Extraction Patterns for Subjective Expressions*. School of Computing University of Utah, 2003 <http://www.cs.utah.edu/~riloff/pdfs/emnlp03.pdf>.

NATO Advanced Study Institute on Recent Advances in Language Engineering for Low- and Middle-Density Languages, and Sergei Nirenburg. *Language engineering for lesser-studied languages*. Amsterdam, Netherlands: IOS Press, 2009.

NetLibrary, Inc, and National Research Council (U.S.). *Technology for the United States Navy and Marine Corps 2000-2035 Becoming a 21st-Century Force*. [Washington, D.C.]: National Academy of Sciences, 1997.

Northam, Jackie. *Pentagon Ramps Up Training to Cut Language Gap*. National Public Radio. February 9, 2007. <http://www.npr.org/templates/story/story.php?storyId=7299421>.

OFFICE OF NAVAL RESEARCH ARLINGTON VA. *Naval S&T Strategic Plan: Defining the Strategic Direction for Tomorrow*. Ft. Belvoir: Defense Technical Information Center, 2009. <http://handle.dtic.mil/100.2/ADA499909>.

OFFICE OF THE UNDER SECRETARY OF DEFENSE FOR ACQUISITION TECHNOLOGY AND LOGISTICS WASHINGTON DC. *Defense Science Board 2006 Summer Study on 21st Century Strategic Technology Vectors, Volume 2: Critical*

Capabilities and Enabling Technologies. Ft. Belvoir: Defense Technical Information Center, 2007. <http://handle.dtic.mil/100.2/ADA464370>.

Office of the Under Secretary of Defense for Acquisition Technology and Logistics Washington DC. *Defense Science Board 2006 Summer Study on 21st Century Strategic Technology Vectors, Volume 2: Critical Capabilities and Enabling Technologies*. Ft. Belvoir: Defense Technical Information Center, 2007.

Ong, Wing S. S. *Sequoyah Foreign Language Translation System - Business Case Analysis*. Ft. Belvoir: Defense Technical Information Center, 2007. <http://handle.dtic.mil/100.2/ADA475922>.

Popp, Robert. *Countering Terrorism Trough Information Technology*. Communications of the ACM. March 2004. Vol. 47/3. <http://information-retrieval.info/taipale/papers/p36-popp.pdf>.

RAND. "Connecting the Dots" in Intelligence. RAND Corporation, Santa Monica, CA, 2005. <http://www.rand.org/pubs/research_briefs/RB9079/RAND_RB9079.pdf>.

Schmidt, Roderic D. *Land Warfare of the Future*. The library of future weaponry. New York: Rosen Pub. Group, 2006.

Schwartz, Mosche. Department of Defense Contractors in Iraq and Afghanistan: Background and Analysis. Congressional Research Service. December 2009. <http://www.fas.org/sgp/crs/natsec/R40764.pdf>.

Shapiro, Amy. *The use of Latent Semantic Analysis as a Tool for the Quantitative Assessment of Understanding and Knowledge.* <https://www.msu.edu/~kidwaikh/cep901b/literaturereview/wee Diekema, Anne R., Wen-Yuan, Hsiao. Cross-*Language Information Retrieval using Dutch Query Translation*. Syracuse University, School of Information Studies, 2001.k04lsa.htm>.

Sherman, Kevin B. *Orchestrating Dominant Battlespace Awareness*. Ft. Belvoir: Defense Technical Information Center, 1998. <http://handle.dtic.mil/100.2/ADA351664>.

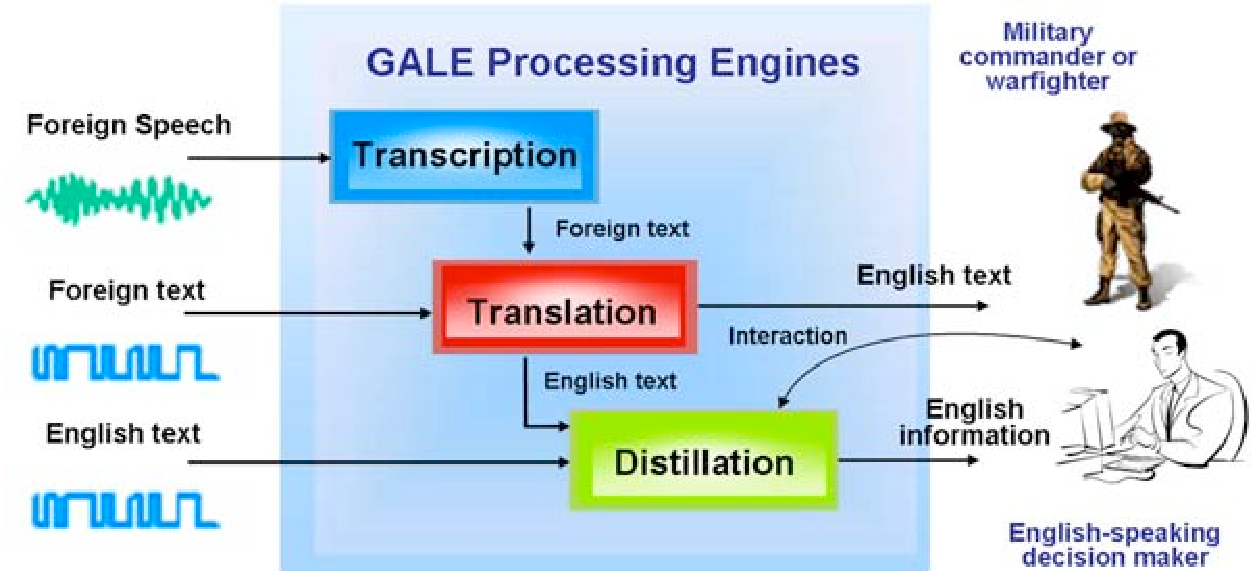
U. S. Army. *Army Vision 2010*. <http://www.army.mil/2010/information_superiority.htm>.

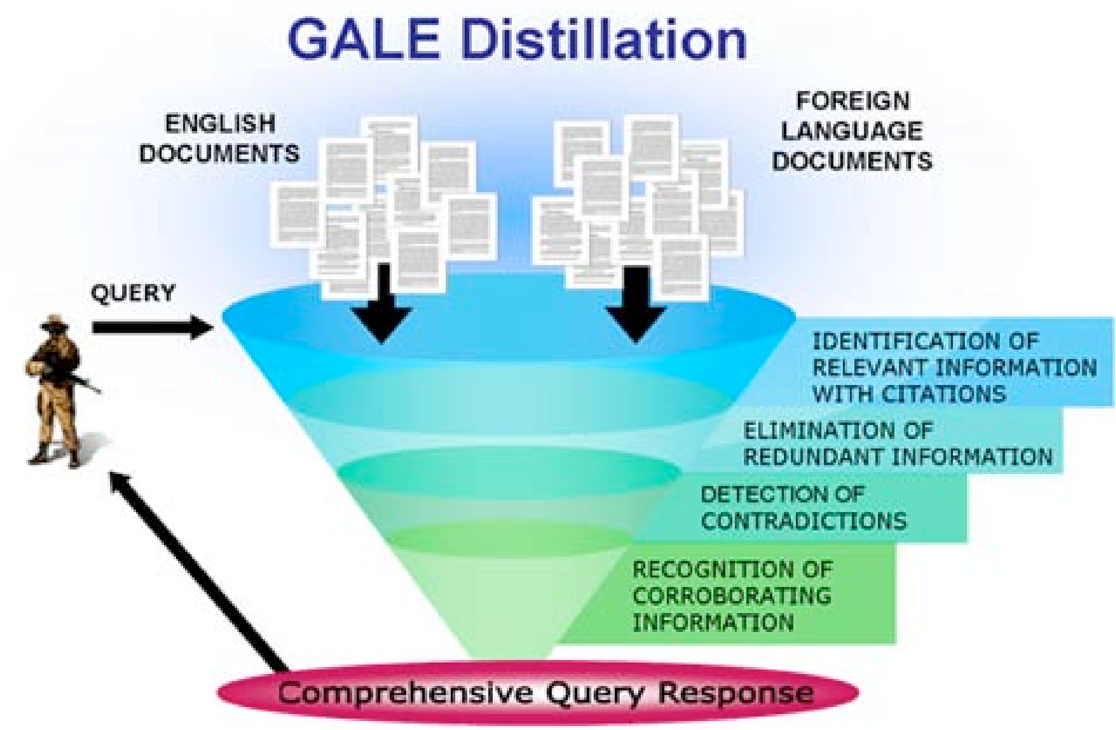
U.S. Department of the Navy, Commandant of the U.S. Marine Corps. Marine Corps Vision and Strategy 2025. PCN 50100654800. Washington D.C.: 2008. <http://www.onr.navy.mil/about/docs/usmc_vision_strategy_2025_0809.pdf>.

Wilks, Yorick. *Machine Translation: Its Scope and Limits*. New York: Springer, 2008.

Appendix A

Global Autonomous Language Exploitation (GALE)





Source: Information Processing Techniques Office. *Global Autonomous Language Exploitation (GALE)*. <http://www.darpa.mil/ipto/Programs/gale/gale_approach.asp>.